

## BIROn - Birkbeck Institutional Research Online

Hu, Y.-T. (2007) Extreme correlation of defaults and LGDs. Working Paper. Birkbeck, University of London, London, UK.

Downloaded from: <https://eprints.bbk.ac.uk/id/eprint/26907/>

*Usage Guidelines:*

Please refer to usage guidelines at <https://eprints.bbk.ac.uk/policies.html>  
contact [lib-eprints@bbk.ac.uk](mailto:lib-eprints@bbk.ac.uk).

or alternatively

ISSN 1745-8587



School of Economics, Mathematics and Statistics

BWPEF 0705

# **Extreme Correlation of Defaults and LGDs**

Yen-Ting Hu  
*Barclays Capital*

February 2007

# **Extreme Correlation of Defaults and LGDs**

Yen-Ting Hu<sup>\*</sup>

First Draft: 28 November, 2004

This Draft: 19 February, 2006

PRELIMINARY

## **Abstract**

This paper conducts a systematic investigation into the correlation between the default rate and three definitions of the recovery rate: price recoveries, settlement recoveries and discounted settlement recoveries. The data suggests a strong linear correlation for price recoveries and a weak one for settlement recoveries, but little or no correlation for discounted settlement recoveries. Using extreme value techniques, I show that the tail dependency for the settlement recoveries is as strong as that for the price recoveries. The probability of high losses (loss given default exceeding 0.9) is consistently higher for the settlement recoveries than for the price recoveries at any level of the quarterly default rate above 0.1%.

<sup>\*</sup> Barclays Capital, Email: [yen-ting.hu@barcap.com](mailto:yen-ting.hu@barcap.com)

I am grateful to William Perraudin, Raymond Brummelhuis and Olivier Renaud for comments and suggestions.

The views expressed are those of the author and not necessarily those of the Barclays Capital.

# 1 Introduction

Loss given default (LGD) is defined as  $1 - \text{recoveries}$ . Due to the complexity of bankruptcy settlement procedures two common practical definitions of recoveries are widely used. The first one, usually termed *price recoveries* uses the price of the obligation at the time of or a short period after default, while the second one, *settlement recoveries*, uses the value of the pre-petition instrument or the settlement instrument when the settlement is reached.

Most research on the cyclical nature of LGDs focuses on a single definition of LGDs. Gupton et al. (2000) document a correlation coefficient of 0.78 for the relationship between LGD (as defined by price recoveries) and default rates using Moody's database on bank loan recoveries between 1989 and 2000. Hu and Perraudin (2002), after standardizing the price recoveries of corporate bonds across different characteristics, show correlations between 0.14 to 0.31 for different sub-samples; they also find an increase in VaR once the correlation is taken into account. The exception is Acharya et al. (2003), who examine the determinants of recoveries using both definitions. Their explanatory variables cover three levels of debt characteristics: firm-specific, industry and macroeconomic. On an industry level, they found that recoveries are 10% to 20% lower when the industry is in distress and this pattern appears in both definitions of recoveries.

Two other papers investigate the reasons for the cyclical nature of the LGDs. Altman et al. (2003) use price recoveries of corporate bonds to show that the weighted average default rate on bonds in the high yield bond market is a statistically significant explanatory variable of the recovery rate. Further, recovery rates are lower in recessions due to an increase in supply of distressed bonds, which is approximated by the total amount of high yield bonds outstanding in the market. Frye (2000a) and (2000b) suggests that, in addition to probability of default (PD), recoveries are also driven by systematic factors through the fluctuation of collateral values. Explicitly building this into the factor model suggested by Finger (1999) and Gordy (2000), Frye tests the model using Moody's dataset and obtained similar estimates of the loading parameters for default and recovery rates.

This paper uses a dataset of defaulted bonds with both price and settlement recoveries and shows that the correlation between defaults and recoveries almost disappears once the settlement recoveries are discounted with the risk-free interest rate. Moreover, a systematic investigation into the cyclical pattern of the two recovery definitions is conducted. Due to the uniqueness of my dataset, I am able to obtain three definitions of recoveries for each bond in the dataset: a) the

price recoveries, b) the settlement recoveries and c) the discounted settlement recoveries. I first examine the cyclical pattern in the correlation between recoveries and default rates. Also, in addition to linear correlation, I employ a bivariate Extreme Value Theory (EVT) model to explore the potential tail dependence when the economy is in the ‘bad’ state. Finally, I investigate the probability of high losses suggested by the bivariate EVT model when the default rate is in different quantiles. The potential implication on capital adequacy is important. Financial institutions that adopt different definitions of recoveries should adjust their capital reserves according to the different cyclical pattern of recoveries.

This paper is organized as follows. Section 2 describes the data and investigates the linear correlation between defaults and recoveries. Section 3 presents the results of EVT analysis and Section 4 investigates the probability of a high default rate and a high LGD happening at the same time. Section 5 concludes.

## 2 The cyclical effect for different definitions of LGDs

### 2.1 Data

The recovery rate is defined as the amount that creditors recover from the defaulted debt for every dollar of debt. Namely, it is defined as the fraction of the money recovered of the face value of the debt.<sup>1</sup> The data I use comes from two sources. The *price recoveries*,  $P_d$ ,<sup>2</sup> were obtained from the Moody’s Corporate Bond Default Database released in July 2002 and the *settlement recoveries*,  $P_e$ ,<sup>3</sup> from Standard and Poor’s Credit Pro 6.4. The Moody’s database consists of details on corporate bonds and issuers that have been rated by Moody’s (including U.S. and non-U.S. domiciled bonds) from January 1970 to January 2002. Price recoveries are defined as the price of defaulted debts 30 days after default. On the other hand, S&P’s database contains default experiences of American public companies covering the period between January 1981 to December 2002. That is, S&P’s database includes all types of outstanding obligations, including bonds, loans etc. of the defaulted companies. The *settlement recoveries* are calculated using three different methods<sup>4</sup>:

---

<sup>1</sup> This is the most widely used definition in the industry, while in the credit pricing model, various other definitions are also used. See Guha (2002) for details.

<sup>2</sup> “ $d$ ” stands for the time of default.

<sup>3</sup> “ $e$ ” stands for the time of emergence.

<sup>4</sup> Credit Pro 6.4 User Guide.

1. Trading price at emergence: the trading prices of pre-petition instruments at the time of emergence,
2. Settlement Pricing: the earliest available trading prices of the instruments received in a settlement, and
3. Liquidity Event Pricing: the value for illiquid settlement instruments at the time of a ‘liquidity event’ – on the first day a price can be determined.<sup>5</sup>

S&P’s provide their chosen methods of recovery definitions for each obligation in the database. In this study, I use their chosen definitions.

To combine the two datasets, I match the CUSIP number<sup>6</sup> of each obligation in the two databases and result in 594 observations that have both  $P_d$  and  $P_e$ . The numbers of observations that use each of the three definitions mentioned above for  $P_e$  are: 145 for “trading price at emergence”, 424 for “settlement pricing” and 25 for “liquidity event pricing”. Table 1 summarizes the numbers of observations per default type in the sample. 87.7% of the observations are defaulted under Chapter 11 or missed principal and/or interest payments, while the rest of the observations are spread in between distressed exchange, pre-packaged Chapter 11 or other less common types of defaults. The average  $P_d$  and  $P_e$  per default type are also reported in Table 1. As observed in Franks and Turous (1994), the average recovery for distressed exchange is higher than that of Chapter 11. This is the case for both price and settlement recoveries. Other types of defaults also seem to exhibit somewhat different average recoveries. However, given the small numbers of total observations after matching and the small proportion of other default types in the sample, all the observations are included in the study despite the default types in order to maximize the sample size.

The descriptive statistics of  $P_d$  and  $P_e$  for the whole sample are shown in Table 2. In addition to the above two recoveries, I also calculated a *discounted settlement recovery*,  $disP_e$ . Unlike some existing research using the coupon rate as a discount factor<sup>7</sup>, I discount the  $P_e$  by the 2 year American government bond yield<sup>8</sup>. I believe that since the coupon rate is decided at the time of issuance, it only reflects the creditworthiness as well as the term to maturity of the bond

---

<sup>5</sup> Examples of the liquidity event pricing are: a company acquisition taken place and the debt is paid off, subsequent distressed exchanges or refinancing, or a callable security is called at the value of the call price etc.

<sup>6</sup> CUSIP stands for Committee on Uniform Securities Identification Procedures. It’s a unique identifier for most securities in the US, including government and corporate bonds and stocks.

<sup>7</sup> For example, Acharya et al. (2003).

<sup>8</sup> The 2-year government bond yield is chosen because the average time a company spends in default in the sample is 1.7 years.

at that time. After the company defaults, the discount factor should reflect the risk of eventual bankruptcy. Since there is no good proxy for this purpose, I chose the risk-free rate to simply capture the time value of money.<sup>9</sup> The descriptive statistics of the discounted settlement recoveries are reported in Table 2.

From Table 2 and the histogram of the three recovery series in Figure 1, it emerges that the settlement recoveries have greater variability than the price recoveries and tend to be concentrated around zero and one. This suggests that in bankruptcy settlements, obligors tend to either suffer complete loss or are able to obtain total recovery. Van de Castle and Keisman (1999) use a similar dataset and obtain a similar distribution.<sup>10</sup> In contrast, price recoveries do not go to zero – for obvious reasons – and are rarely equal to one or more. The smaller volatility of the price recoveries is in line with the general view that price recovery is the expected discounted settlement recovery. Moreover, the mean of the  $disP_e$  being 15% higher than the mean of the  $P_d$  could also reflect a risk premium for holding defaulted bonds, and the returns for investors such as “vulture funds” who make profits from buying and selling distressed debt.

## 2.2 The Cyclical pattern for different recovery definitions

### 2.2.1 Linear correlation

To examine the relationship between defaults and LGDs,<sup>11</sup> I look at them at the aggregate level. That is, I calculate the average recovery rate per quarter from Q1 1984 to Q4 2002 according to either the time of default for  $P_d$  and  $disP_e$  or the time of settlement for  $P_e$ . Each of the series is then transformed to  $LGD$ , where  $LGD(P_d) = 1 - P_d$ ,  $LGD(disP_e) = 1 - disP_e$  and  $LGD(P_e) = 1 - P_e$ .

As for the default rate, I calculate the number of defaults in Moody’s database in a given quarter, divided by the total number of outstanding bonds rated by Moody’s. This results in 76 quarterly observations. The descriptive statistics for the default rate are shown in Table 2. After obtaining the quarterly default rate, I pair them up with the three quarterly LGD series. Notice that the  $LGD(P_d)$  and  $LGD(disP_e)$  are paired up with the default rate at the time of default, while

---

<sup>9</sup> A recent paper by Maclachlan (2004) proposes a framework based on CAPM for choosing the appropriate discount rate for different asset classes, such as defaulted corporate bonds and bank loans.

<sup>10</sup> Davydenko, S. and Franks, J. (2005) report highly skewed distributions similar distributions for small firm bank loan defaults in UK and Germany, while the distribution for France is bimodal.

<sup>11</sup> I prefer LGD to recoveries for all subsequent analysis because it’s easier to express in terms of large losses than small recoveries are. It also makes EVT analysis easier.

$LGD(P_e)$  is paired up with the default rate at the time of settlement. The purpose is to see the role of the default rate at the time when the market evaluates the bonds.

Due to the loss of observations in the process of matching price and settlement recoveries information, the numbers of quarterly observations for  $LGD(P_d)$ ,  $LGD(disP_e)$  and  $LGD(P_e)$  are less than that of the default rate series (59, 59 and 58 respectively). Figure 2 to Figure 4 show the scatter plots for the default rate and the three definitions of LGDs. In each figure, I also fit a linear regression of LGDs with the default rate as the explanatory variable. The regression suggests that while the default rate might have some explanatory power over  $LGD(P_d)$  and  $LGD(P_e)$ , the slope for the  $DR-LGD(disP_e)$  case is virtually zero.<sup>12</sup> In addition, while Figure 2 suggests that the correlation between default rate and  $LGD(P_d)$  can probably be approximated by a linear relationship, Figure 4 actually suggests a possibly non-linear relationship between default rate and  $LGD(P_e)$ .

Table 3 reports the linear correlation for the three pairs of observations. The numbers in brackets report the standard error for the respective correlation coefficients. One can see that the correlation is strongest for the  $DR-LGD(P_d)$  pair, and then for the  $DR-LGD(P_e)$  pair. For the  $DR-LGD(disP_e)$  pair, the correlation is not statistically significantly different from zero according to the t-statistics.

From a risk management point of view, the most interesting case is the time when there are many defaults. In the next section, I examine the relationship between default and recovery rates when the default rate is high.

### 2.2.2 Conditional average LGD and correlations

In order to examine the behaviour of LGDs in stressed periods, namely when the default rate is high, I introduce two new variables; conditional average LGD ( $ALGD$ ) and conditional correlations. These are calculated in the following way. For pairs of  $DR$  and  $LGD$ , I order the data according to  $DR$ . Suppose  $(DR_k, LGD_k)$  denotes the  $k^{th}$  largest  $DR$  and its associated  $LGD$ . Then the  $ALGD$  is defined as the average  $LGD$  where  $k$  equals from 1 to  $K-1$ , namely those LGDs that are associated with default rates larger than  $DR_k$ . I also define conditional correlation as the correlation between the  $DR_k$  and  $LGD_k$  for all  $k$  between 1 and  $K-1$ .

---

<sup>12</sup> Altman et al. (2001) reports that the face-value-weighted average default rate is one of the significant explanatory variable for the market-value-weighted average price recovery rate.



I first plot the *ALGD* against the default rate percentile in Figure 5. The left hand most observation gives the average LGD for all observations, which is also the unconditional average LGD for the whole sample. As one moves further to the right, the  $y$  level of the line shows the conditional mean of the LGDs subject to their associated DRs being greater than the  $x$  percentile of the default rate distribution.

Even though the right hand side of the figure suffers from high volatility due to the small amount of observations, it still suggests an increasing level of average LGD up to around 80% default rate percentile if one conditions on progressively worse stress states as reflected in default rate levels. This is the case for all three definitions of LGD; however, the extent of the increase is not the same. The rises of ALGD from the unconditional one to the one beyond the 80% default rate percentile, for example, are 17.58% and 18.54% for ALGD defined by  $P_d$  and  $P_e$ , while the rise of ALGD defined by discounted  $P_e$  is only 6.54%.

The same idea is used to plot the conditional correlation in Figure 6. The correlations stay almost constant up to around 70% default rate percentile. As in the case of conditional average LGD, the right hand side of the plot suffers from exceptional volatility due to lack of observations. This provides an incentive to study the dependence structure in the tail using Extreme Value Theory.

To sum up, the correlation analysis suggests that the dependency between the default and recovery rates only exists when the timing of the default rate and the recoveries measurement coincide. Furthermore, this dependency contributes to over 15% of the increase in the average LGD in stressed periods. In the next section, I will focus on the dependence structure between default rates and LGDs at the tail of the distribution.

### 3 EVT analysis on tail dependence

#### 3.1 Methodology

##### 3.1.1 Asymptotic Results<sup>13</sup>

Suppose that  $(X_1, Y_1), (X_2, Y_2), \dots$  is a sequence of vectors that are randomly drawn from a distribution  $F(x, y)$ . If the marginal distributions,  $F_x$  and  $F_y$ , are transformed into unit Fréchet distributions, with distribution function

---

<sup>13</sup> See Coles (2001) for a detail derivation.

$$F(z) = \exp(-1/z), \quad z > 0 \quad (1)$$

then there exists a non-degenerate distribution  $G(\tilde{x}, \tilde{y})$  of the transformed variables  $\tilde{x}$  and  $\tilde{y}$  which, in the limit, takes the form

$$G(\tilde{x}, \tilde{y}) = \exp\{-V(\tilde{x}, \tilde{y})\}, \quad \tilde{x} > 0, \tilde{y} > 0 \quad (2)$$

where

$$V(\tilde{x}, \tilde{y}) = 2 \int_0^1 \max\left(\frac{w}{\tilde{x}}, \frac{1-w}{\tilde{y}}\right) dH(w),$$

and  $H$  is a distribution function on  $[0, 1]$  satisfying the mean constraint

$$\int_0^1 w dH(w) = 1/2.$$

Here  $V$  is the dependence function. The complication of the above result is that the multivariate distribution is not completely specified since the shape of the dependence function is unknown. One standard class of functions that satisfies the above constraint for  $H$  is the logistic family. In this case

$$G(\tilde{x}, \tilde{y}) = \exp\left\{-\left(\tilde{x}^{-1/\alpha} + \tilde{y}^{-1/\alpha}\right)^\alpha\right\}, \quad \tilde{x} > 0, \tilde{y} > 0 \quad (3)$$

for a parameter  $\alpha \in (0, 1)$ , which is the measure of dependency. When  $\alpha = 1$ , it represents asymptotic independence, and  $\alpha = 0$  for total dependence.

### 3.1.2 Threshold Excess Model

With the threshold excess model, the tail of a distribution takes the form of the General Pareto distribution

$$G(x) = \begin{cases} 1 - \left\{1 + \frac{\xi(x-u)}{\sigma}\right\}^{-1/\xi}, & x > u, \xi \neq 0 \\ 1 - \exp\left\{-\frac{(x-u)}{\sigma}\right\}, & x > u, \xi = 0 \end{cases} \quad (4)$$

where  $x \geq 0$  when  $\xi \geq 0$  and  $0 \leq x \leq -\sigma/\xi$  when  $\xi < 0$ . This distribution has three parameters;  $u$  is the *position* parameter, i.e. the threshold value,  $\xi$  is the *shape* parameter, and  $\sigma$  is an additional scaling parameter. The asymptotic theory of EVT proves that with high enough  $u$ ,  $F(x) \approx G(x)$  on  $x > u$ .<sup>14</sup>

---

<sup>14</sup> See Embrechts et al. (2003) for a detail derivation.

For suitable thresholds  $u_x$  and  $u_y$ , the marginal distributions of  $F$  each have an approximation form as equation (4), with respective parameter sets  $(\lambda_x, \sigma_x, \xi_x)$  and  $(\lambda_y, \sigma_y, \xi_y)$ , where  $\lambda$  is the probability of the observations exceeding the threshold. The transformations would then be

$$\tilde{X} = - \left\{ \log \left[ 1 - \lambda_x \left( 1 + \frac{\xi_x (X - u_x)}{\sigma_x} \right)^{-1/\xi_x} \right] \right\}^{-1} \quad (5)$$

and

$$\tilde{Y} = - \left\{ \log \left[ 1 - \lambda_y \left( 1 + \frac{\xi_y (Y - u_y)}{\sigma_y} \right)^{-1/\xi_y} \right] \right\}^{-1} \quad (6)$$

Applying the asymptotic result above, the joint distribution of the transformed variables is:

$$\begin{aligned} \tilde{F}(\tilde{x}, \tilde{y}) &= \{\tilde{F}^n(\tilde{x}, \tilde{y})\}^{1/n} \\ &\approx [\exp\{-V(\tilde{x}/n, \tilde{y}/n)\}]^{1/n} \\ &= \exp\{-V(\tilde{x}, \tilde{y})\} \end{aligned}$$

Finally, since  $F(x, y) = \tilde{F}(\tilde{x}, \tilde{y})$ , it follows that

$$F(x, y) \approx G(x, y) = \exp[-V(\tilde{x}, \tilde{y})] \quad (7)$$

The estimation is simplified by the fact that one can consider  $X_i$  and  $Y_i$  separately as independent univariate random variables, and obtain the estimates of  $(\sigma_x, \xi_x)$  and  $(\sigma_y, \xi_y)$  separately.  $\lambda$  can be estimated by the empirical probability, i.e. the number of exceedances divided by the number of observations.

In order to utilize all the available data points, especially in a small sample, various techniques can be applied to deal with the missing observation from the pair of data. One way that is being adopted here is to use the extended data to estimate the marginal distribution and then use these estimates to transform the bivariate data, i.e. data without missing values, prior to estimation of the joint distribution.<sup>15</sup>

The likelihood estimator of the joint distribution is complicated by the fact that possibly only one of the pair of observations exceeds the threshold values. When this is the case, only the

---

<sup>15</sup> Another way is to estimate the parameter for the univariate and bivariate models simultaneously. For the pair of observations where only one of them is observable, only the marginal distribution of the observed one would enter the likelihood function.

partial effect of the one that exceeds the threshold should contribute to the likelihood. Therefore, the likelihood function is as follows:

$$L(\theta; (x_1, y_1), \dots, (x_n, y_n)) = \prod_{i=1}^n \psi(\theta; (x_i, y_i)), \quad (8)$$

where  $\theta$  denotes the parameters of  $F$  and

$$\psi(\theta; (x, y)) = \begin{cases} \left. \frac{\partial^2 F}{\partial x \partial y} \right|_{(x,y)}, & \text{if } x > u_x \text{ and } y > u_y, \\ \left. \frac{\partial F}{\partial x} \right|_{(x, u_y)}, & \text{if } x > u_x \text{ and } y \leq u_y, \\ \left. \frac{\partial F}{\partial y} \right|_{(u_x, y)}, & \text{if } x \leq u_x \text{ and } y > u_y, \\ F(u_x, u_y), & \text{if } x \leq u_x \text{ and } y \leq u_y, \end{cases}$$

with each term being derived from equation (7).<sup>16</sup>

### 3.1.3 Threshold selection

In practice, the application of the asymptotic results for univariate threshold excess models depends crucially on the selection of threshold values. If the value is too high, it results in inefficient parameter estimates because there are too few observations used for estimation. On the other hand, if the threshold value is too low, one would be using observations that are not in the tail for the estimation, which leads to biased estimates. To optimize this trade off between bias and inefficiency, Login and Solnik (2001) suggested a Monte Carlo approach, where the optimal value is chosen on the basis of the lowest mean square error (MSE).

The MSE is calculated as follows. Suppose  $X_i$ , where  $i = 1$  to  $n$ , is the estimates obtained from  $N$  simulations that are drawn from a chosen distribution, which is the assumed parent distribution of the underlying data, and  $\theta$  is the parameter value given the chosen distribution, the MSE can be decomposed as follows:

$$\text{MSE} = (\bar{X} - \theta)^2 + \frac{1}{N} \sum_{i=1}^N (X_i - \theta)^2 \quad (9)$$

where  $\bar{X}$  represents the mean of  $N$  simulated estimations. The first part of the decomposition measures the bias and the second part the inefficiency. Jansen and de Vries (1991) have shown that there is a U-shaped relationship between the MSE and the numbers of observations

---

<sup>16</sup> See Appendix 1 for the derivation.

exceeding the threshold. The minimum MSE point can then be used for the estimation for the model.

Suppose  $\kappa$  is the number of excesses where the MSE is minimal in the simulation exercise, one can then estimate the univariate EVT distribution using the real data with numbers of excesses from 1 to  $\kappa$ . The optimal threshold value and the estimation of the parameters are then decided simultaneously. The optimal number of exceedances,  $n$ , is selected for which the estimated index is statistically closest to the theoretical value of the tail index defined in the simulation procedure. Practically speaking, one considers the  $p$  value of the  $t$  test for the hypothesis:  $\zeta(n) = \theta$  and chooses the threshold value,  $u$ , where the  $p$  value is highest.

### 3.2 Estimation Results

#### 3.2.1 Monte Carlo threshold selection

The random draw series is generated under the assumption that the underlying data is Beta distributed with two shape parameters  $\alpha$  and  $\beta$  and density function:

$$f(x : \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad 0 \leq x \leq 1$$

where  $B(\alpha, \beta)$  is the beta function, and  $\alpha, \beta > 0$ . There are two reasons for this choice. First, both the default rate and the LGDs lie between 0 and 1. There are two observations in the settlement recoveries series  $LGD(P_e)$  that exceed 1; I consider them to be outliers and replace them with 1. Second, the two-parameter Beta distribution allows great flexibility on the shape of the distribution, which would, in turn, provide more robust results for the threshold selection.

The estimators are obtained by setting the first and second moments of the beta distribution equal to the sample mean and variance:

$$\begin{aligned} \alpha^* &= \bar{x} \left( \frac{\bar{x}(1-\bar{x})}{s^2} - 1 \right) \\ \beta^* &= (1-\bar{x}) \left( \frac{\bar{x}(1-\bar{x})}{s^2} - 1 \right) \end{aligned} \tag{11}$$

where  $\bar{x}$  is the sample mean and  $s^2$  is the sample variance. Table 5 shows estimation results for the four data series. The exceptionally large  $\beta^*$  for the default rate series suggests that the beta distribution is skewed to the left. This is the case since empirical default rates tend to be very small and never come close to 1, especially considering that I'm using quarterly default rates.

Using the parameters estimated above, I simulated 1000 replications of 76, 59, 58 and 58 draws for the data series of  $DR$ ,  $LGD(P_d)$ ,  $LGD(disP_e)$  and  $LGD(P_e)$  respectively. The Mean Squared Errors (MSE) of the tail index  $\xi$  with 5 to 50 exceedances are calculated and shown in Figure 7. The theoretical value of  $\xi$  is related to the parent beta distribution by  $\xi = -1/\beta$ .<sup>17</sup> As suggested by Jansen and de Vries (1991), the MSE exhibited a U shape with the increase in exceedances. Table 6 shows the numbers of exceedances where the MSE is minimal.

### 3.2.2 Marginal EVT estimation

At this point, the marginal EVT distributions for  $DR$ ,  $LGD(P_d)$ ,  $LGD(disP_e)$  and  $LGD(P_e)$  are estimated with the number of exceedances from 5 to the level where the minimum MSE is found as shown in Table 6. The optimal threshold value is decided by the  $p$  value of the  $t$ -test of the hypothesis that the estimated tail index is equal to the theoretical value of the tail index given by the beta distribution. In the meantime, the estimation of the parameters is also determined accordingly. The result is shown in Table 7.

### 3.2.3 Bivariate EVT estimation

After transforming the data with the estimates above, one can estimate the dependence level based on the assumption of a logistic dependence function. Note that in the bivariate estimation, only those observations that are available to all four series are used. This leaves only 51 observations in total. The numbers of exceedances are then reduced to 23, 11, 11 and 12 for default rate,  $LGD(P_d)$ ,  $LGD(disP_e)$  and  $LGD(P_e)$ , respectively. The results are shown in Table 8. The  $\alpha$  parameter, which lies between 0 and 1, measures the dependence level between two variables. The smaller the  $\alpha$  is, the bigger the dependency, and it is related to the linear correlation coefficient by  $\rho = 1 - \alpha^2$  (Tiago de Oliveira, (1973)). One can see that in the tail, there is still not much correlation between defaults and  $LGD(disP_e)$ , although it is bigger than the linear correlation. The interesting result is that the dependence level for the  $DR$ - $LGD(P_e)$  pair is around the same level as for the  $DR$ - $LGD(P_d)$  pair. This suggests that while in the ‘normal’ state of the economy the correlation between defaults and  $LGD(P_e)$  is not particularly high, the relationship is strengthened in the ‘bad’ state of the economy.<sup>18</sup> In the next section, I investigate the implication of this relationship.

---

<sup>17</sup> See Embrechts (2003) for a detailed derivation.

<sup>18</sup> Similar pattern is observed when the rank correlations are being calculated for the dataset.

## 4 Probability of large losses given a high default rate

To see how the dependence structure suggested by EVT affects risk management decisions, I calculated the probability of large losses given that the default rate is higher than a certain level in the tail. Namely,

$$\text{Prob}(LGD > l \mid DR > d), \text{ for } l > u_{LGD}, d > u_{DR} \quad (12)$$

where  $u_{LGD}$  and  $u_{DR}$  are threshold values obtained in the univariate EVT estimation. This probability can be calculated using the marginal and the joint distribution function with the parameters estimated in section 3.2.2 and section 3.2.3.

Suppose  $F_{LGD}$  and  $F_{DR}$  are the marginal EVT distributions for  $DR$  and  $LGD$ s, and  $F_{LGD,DR}$  is their joint EVT distribution. Then,

$$\begin{aligned} & \text{Prob}(LGD > l \mid DR > d), \text{ for } l > u_{LGD}, d > u_{DR} \\ &= \frac{\text{Prob}(LGD > l, DR > d)}{\text{Prob}(DR > d)} \\ &= \frac{1 - [F_{LGD}(l) + F_{DR}(d) - F_{LGD,DR}(l, d)]}{1 - F_{DR}(d)} \end{aligned}$$

I first generate a series of quarterly default rates from its threshold value 0.07% to a sufficiently large number, 1.5%. I then calculate the above conditional probability when  $d$  equals from 0.07% to 1.5%. Figure 8 and Figure 9 show the conditional probability of LGD larger than 0.8 and 0.9.

In both cases, the probability of large losses increases with the level of the default rate. Also, given the same level of default rate, the conditional probability is consistently higher for the  $LGD(P_e)$  than for the  $LGD(P_d)$ . The slope of the increase is smallest for the  $LGD(disP_e)$  among the three definitions.

## 5 Conclusion

In this paper, I systematically examine the dependence structure between the default rate and three different definitions of recoveries. Under the linear, Gaussian assumptions, the default rate at the time of default exhibits strong correlation with LGD as defined by price recoveries, and the default rate at the time of settlement has a rather weak correlation with LGD as defined

by settlement recoveries. Meanwhile, the correlation between default rate and LGD defined by discounted settlement recoveries is not statistically significantly different from zero. Further study can be suggested here to explore the reason for the lack of correlation.

I have also shown that the normal, linear correlation assumption limits us from exploring the dependence structure at the tail of the joint defaults – LGDs distribution. This is what matters most in risk management – the potential large losses given that the economy is in the ‘bad’ state. I employ bivariate Extreme Value Theory models on the three pairs of data,  $DR-LGD(P_d)$ ,  $DR-LGD(disP_e)$  and  $DR-LGD(P_e)$ . The interesting finding is that the dependency between defaults and  $LGD(P_e)$  is a lot stronger in the tail than the linear correlation suggests. In fact, it is at the same level as the correlation for  $DR-LGD(P_d)$ . The reason for this increase might be that the market requires a higher risk premium in high default times.

To see the extent of this dependency on the potential losses, I calculate the probability of high losses, i.e.  $LGD > 0.8$  and  $LGD > 0.9$ , given that the quarterly default rate is above a certain value ranging from 0.7% to 1.5%. I found that the conditional probability of large losses increases with the default rate level. Furthermore, given the same level of the default rate, the conditional probability of large losses is consistently higher for  $LGD(P_d)$  than for  $LGD(P_e)$ . This result is likely to have important implications for risk management. Obviously, given that the probability of large losses is high when the default rate is high, banks need to maintain higher reserves against potential credit losses when the economy is in a downturn than any standard credit risk model would dictate.

These results also suggest different loss measures for different institutions. A mutual fund manager for example, who sells off his/her bond position as soon as it defaults, is likely to use LGD as derived from price recoveries to assess the fund’s potential loss. For banks on the other hand, who have to hold their loan position until the time of settlement, LGD as derived from the settlement recovery would be the more appropriate yardstick. This implies that banks require credit risk models that are more sensitive to the state of the economy than mutual funds.



## Appendix 1 : Likelihood function for bivariate EVT distribution estimation with logistic dependence variable

Suppose  $F$  is the joint distribution of  $x$  and  $y$  on regions of  $x > u_x, y > u_y$ , where  $u_x$  and  $u_y$  are high enough thresholds for  $x$  and  $y$ . With a logistic dependence function,

$$F(x, y) = \exp\{-(x^{-1/\alpha} + y^{-1/\alpha})\} \quad (\text{A2.1})$$

then

$$\begin{aligned} \partial F / \partial x &= \exp\{-(x^{-1/\alpha} + y^{-1/\alpha})\}(-\alpha)(x^{-1/\alpha} + y^{-1/\alpha})^{\alpha-1}(-1/\alpha)x^{-1/\alpha-1} \\ &= \exp\{-(x^{-1/\alpha} + y^{-1/\alpha})\}(x^{-1/\alpha} + y^{-1/\alpha})^{\alpha-1}x^{-1/\alpha-1} \end{aligned}$$

$$\begin{aligned} \partial F / \partial y &= \exp\{-(x^{-1/\alpha} + y^{-1/\alpha})\}(-\alpha)(x^{-1/\alpha} + y^{-1/\alpha})^{\alpha-1}(-1/\alpha)y^{-1/\alpha-1} \\ &= \exp\{-(x^{-1/\alpha} + y^{-1/\alpha})\}(x^{-1/\alpha} + y^{-1/\alpha})^{\alpha-1}y^{-1/\alpha-1} \end{aligned}$$

$$\begin{aligned} \partial^2 F / \partial x \partial y &= \frac{\partial F}{\partial y} \left[ (x^{-1/\alpha} + y^{-1/\alpha})^{\alpha-1} x^{-1/\alpha-1} \right] \\ &\quad + \exp\{-(x^{-1/\alpha} + y^{-1/\alpha})\} x^{-1/\alpha-1} (\alpha-1)(x^{-1/\alpha} + y^{-1/\alpha})^{\alpha-2} \left( \frac{-1}{\alpha} y^{-1/\alpha-1} \right) \\ &= \exp\{-(x^{-1/\alpha} + y^{-1/\alpha})\} \\ &\quad \left[ (x^{-1/\alpha} + y^{-1/\alpha})^{2(\alpha-1)} (xy)^{-1/\alpha-1} + \frac{1-\alpha}{\alpha} (xy)^{-1/\alpha-1} (x^{-1/\alpha} + y^{-1/\alpha})^{\alpha-2} \right] \\ &= \exp\{-(x^{-1/\alpha} + y^{-1/\alpha})\} (xy)^{-1/\alpha-1} \\ &\quad \left[ (x^{-1/\alpha} + y^{-1/\alpha})^{2(\alpha-1)} + (\alpha^{-1} - 1)(x^{-1/\alpha} + y^{-1/\alpha})^{\alpha-2} \right] \end{aligned}$$

## Bibliography

- [1] Acharya, V., S. Bharath, and A. Srinivasan (2003) “Understanding the recovery rates on defaulted securities”, London Business School, September.
- [2] Allen, L. and A. Saunders (2003) “A survey of cyclical effects in credit risk measurement models”, *BIS Working Papers*, No. 126, January.
- [3] Altman, E., A. Resti and A. Sironi (2001) “Analyzing and Explaining Default Recovery Rates” *A Report submitted to The International Swaps & Derivatives Association*, December.
- [4] Altman, E., B. Brady, A. Resti and A. Sironi (2003) “The link between default and recovery rates: Theory, Empirical Evidence and Implications”, *mimeo*, March.
- [5] Altman, E. I. and V. Kishore (1996) “Almost Everything you Wanted to Know About Recoveries on Defaulted Bonds,” *Financial Analysts Journal*, November-December, 57-64.
- [6] Coles, S. (2001) *An Introduction to Statistical Modeling of Extreme Values*, Springer, London.
- [7] Davydenko, S. A. and J. R. Franks (2005) “Do Bankruptcy Codes matter? A Study of defaults in France, Germany and the UK”, London Business School, August.
- [8] Embrechts, P., C. Kluppelberg and T. Mikosch (2003) *Modelling Extremal Events for Insurance and Finance*, Springer, New York.
- [9] Embrechts, P. (ed) (2002) *Extremes and integrated risk management*, Risk Books, London.
- [10] Finger, C. (1999) “Conditional approaches for CreditMetrics portfolio distributions”, *CreditMetrics Monitor*, April.
- [11] Finger, C. (2002) *Creditgrades – Technical Document*, The RiskMetrics Group, New York.
- [12] Franks, J. R. and W. N. Torous (1994) “A comparison of financial recontracting in distressed exchanges and Chapter 11 reorganizations”, *Journal of Financial Economics*, 35, 349-70.
- [13] Frye, J. (2000a) “Collateral damage”, *Risk*, April.

- [14] Frye, J. (2000b) “Depressing Recoveries,” *Risk*, November, 108-11.
- [15] Gordy, M. (2000) “A comparative anatomy of credit risk models”, *Journal of Banking and Finance*, pp. 119 – 49, January.
- [16] Guha, R. (2002) “Recovery of Face Value at Default: Theory and Empirical Evidence” London Business School.
- [17] Gupton, G. M., D. Gates and L. V. Carty (2000) “Bank-Loan Loss Given Default”, Moody’s Investors Service, *Global Credit Research*, November.
- [18] Hu, Y. and W. Perraudin (2002) “The dependence of recovery rates and defaults”, *Bank of England mimeo*, January.
- [19] Jasen, D. W. and C. G. de Vries (1991) ‘On the frequency of large stock returns: putting booms and busts into perspectives’, *Review of Economics and Statistics*, Vol. 73, pp. 18-24.
- [20] Joe, H (1997) *Multivariate Models and Dependence Concepts*, Chapman & Hall, London.
- [21] Johnson, N. L. and S. Kotz (1994) *Continuous Univariate Distributions*, 2<sup>nd</sup> edn., John Wiley, New York.
- [22] Longin, F. and B Solnik (2001) ‘Extreme Correlation of International Equity Markets’, *Journal of Finance*, Vol. LVI, No. 2, pp. 649-76, April.
- [23] Maclachlan, I. (2004) “Choosing the Discount Factor for Estimating Economic LGD”, unpublished, [http://www.defaultrisk.com/pp\\_recov\\_18.htm](http://www.defaultrisk.com/pp_recov_18.htm).
- [24] Schuermann, T. (2003) “What do we know about Loss-Given-Default?” Federal Reserve Bank of New York, March.
- [25] Tiago de Oliveira, J. (1973) “Statistical extremes – A Survey”, Center of Applied Mathematics, Lisbon.
- [26] Van de Castle, K. and D. Keisman (1999) “Recovering Your Money: Insights Into Losses From Defaults”, *Standard and Poor’s CreditWeek*, June.

## Tables

**Table 1 Numbers of observations and average recoveries per default type in the sample**

Default Type	No. of Observations	Average Recoveries		
		$P_d$	$disP$	$P_e$
Chapter 11	238	0.330	0.336	0.386
Distressed exchange	25	0.429	0.382	0.444
Grace period default	11	0.394	0.982	0.999
Missed interest payment	276	0.322	0.377	0.437
Missed principal and interest payments	7	0.503	0.640	0.666
Prepackaged Chapter 11	31	0.395	0.529	0.546
Suspension of payments	6	0.368	0.887	0.944

**Table 2 Descriptive statistics for recovery and default rates**

	$P_d$	$disP_e$	$P_e$	$DR$
Mean	0.337	0.384	0.441	0.0018
Standard Deviation	0.236	0.362	0.428	0.0017
Minimum	0.004	0.000	0.000	0.0001
Maximum	1.018	1.715	1.722	0.0069
No. of Observations *	594	594	594	76

\* for  $P_d$ ,  $disP_e$  and  $P_e$  series, numbers of observations are numbers of bonds in the sample, while for DR, it is the numbers of quarterly observations.

**Table 3 Linear Correlations***no. of quarterly observations without missing values: 51*

	<i>DR</i>	<i>LGD(P<sub>d</sub>)</i>	<i>LGD(disP<sub>e</sub>)</i>	<i>LGD(P<sub>e</sub>)</i>
<i>DR</i>	1.000	0.441 (0.128)	0.069 (0.143)	0.238 (0.139)
<i>LGD(P<sub>d</sub>)</i>	0.441	1.000	0.464	0.342
<i>LGD(disP<sub>e</sub>)</i>	0.069	0.464	1.000	0.110
<i>LGD(P<sub>e</sub>)</i>	0.238	0.342	0.110	1.000

**Table 4 Rank Correlations***No. of quarterly observations: 51.*

	<i>LGD(P<sub>d</sub>)</i>	<i>LGD(disP<sub>e</sub>)</i>	<i>LGD(P<sub>e</sub>)</i>
<i>DR</i>	0.399	0.047	0.308

**Table 5 Parameters for beta distributions**

	$\alpha^*$	$\beta^*$	<i>No. of obs.</i>
<i>Default Rate</i>	1.064	587.544	76
<i>LGD(P<sub>d</sub>)</i>	7.988	4.588	59
<i>LGD(disP<sub>e</sub>)</i>	2.493	1.443	59
<i>LGD(P<sub>e</sub>)</i>	1.635	1.285	58

**Table 6 Minimum Mean Squared Errors**

	<i>Default Rate</i>	<i>LGD(P<sub>d</sub>)</i>	<i>LGD(disP<sub>e</sub>)</i>	<i>LGD(P<sub>e</sub>)</i>
<i>No. of excesses</i>	50	28	23	20
<i>Bias squared</i>	0.179	0.238	0.086	0.054
<i>Variance</i>	0.036	0.183	0.065	0.035
<i>MSE</i>	0.215	0.421	0.151	0.089

**Table 7 Maximum P-values for all bonds**

	<i>Default Rate</i>	<i>LGD(<math>P_d</math>)</i>	<i>LGD(dis<math>P_d</math>)</i>	<i>LGD(<math>P_d</math>)</i>
$n^*$	50	11	16	15
$u^*$	0.001	0.748	0.768	0.724
$\xi^*$	-0.084	-0.228	-0.611	-0.707
$\sigma^*$	0.002	0.060	0.148	0.159
$P(\xi^* = -1/\beta^*)$	0.764	0.989	0.869	0.883

**Table 8 Dependence Level between Default Rates and LGDs**

	<i>LGD(<math>P_d</math>)</i>	<i>LGD(dis<math>P_d</math>)</i>	<i>LGD(<math>P_d</math>)</i>
$\rho$	0.327	0.135	0.363
$\alpha$	0.820	0.930	0.798
<b>S.E. (<math>\alpha</math>)</b>	0.080	0.079	0.087

## Figures

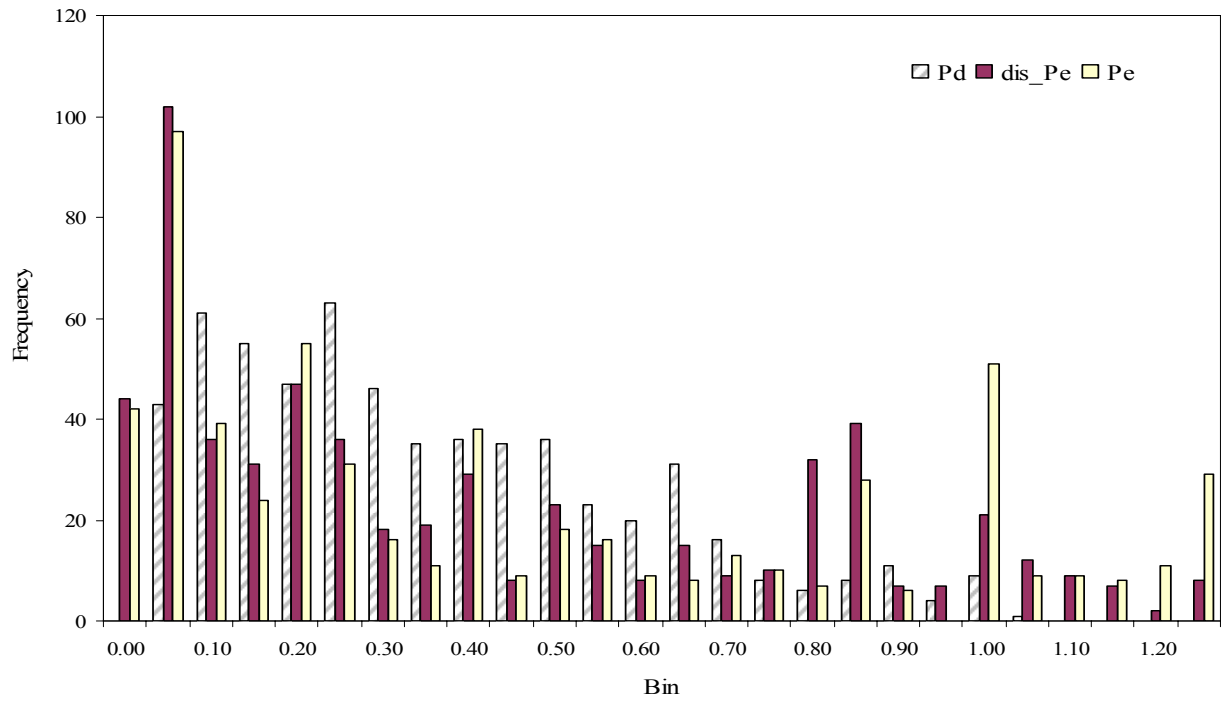


Figure 1 Histogram of  $P_e$ ,  $P_d$ , and discounted  $P_e$

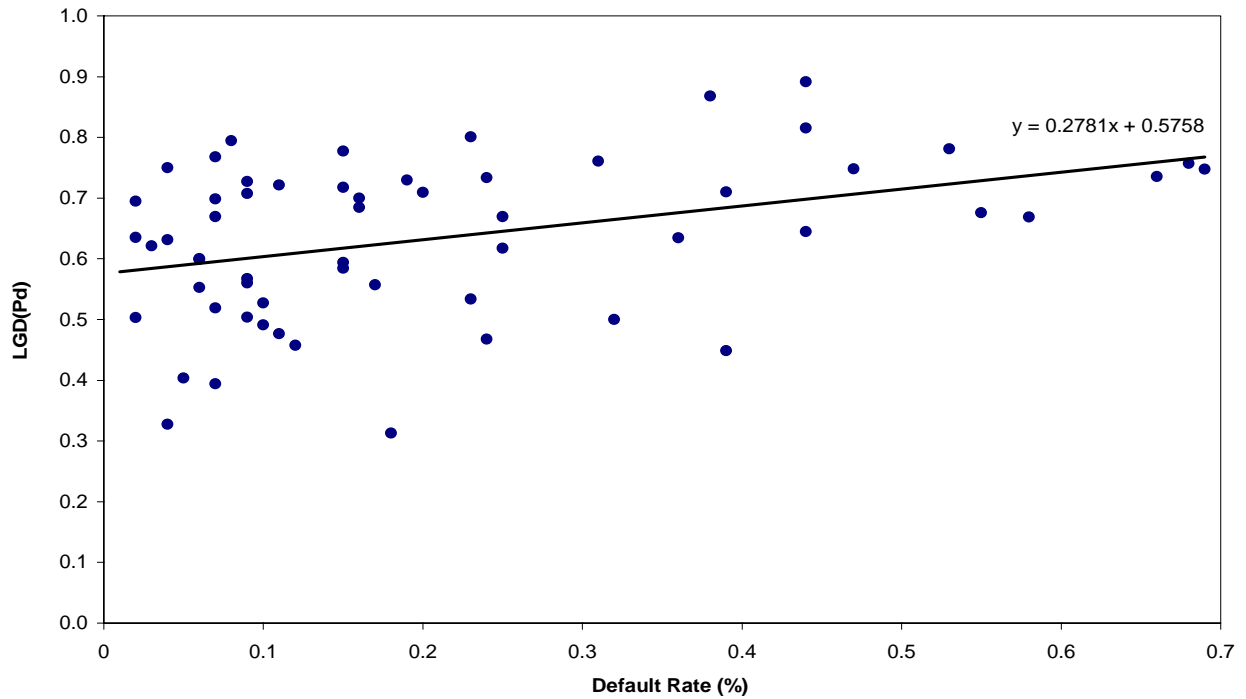
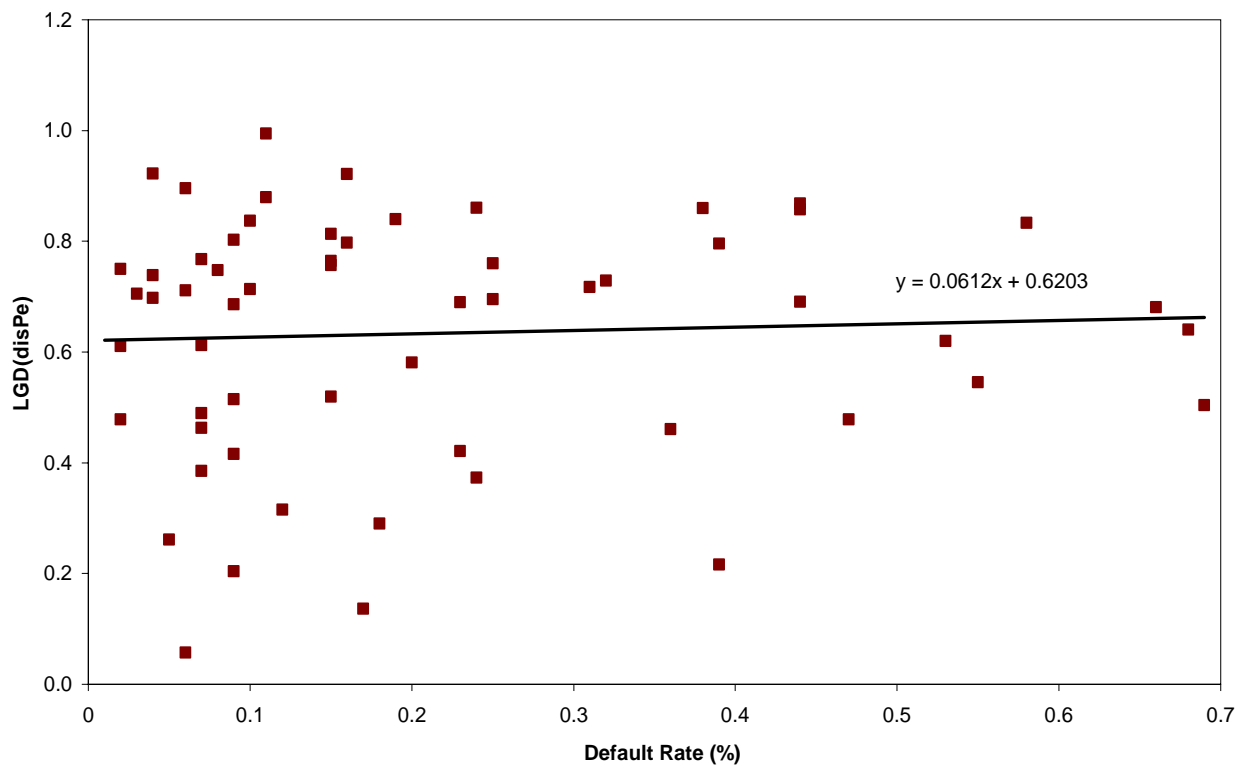
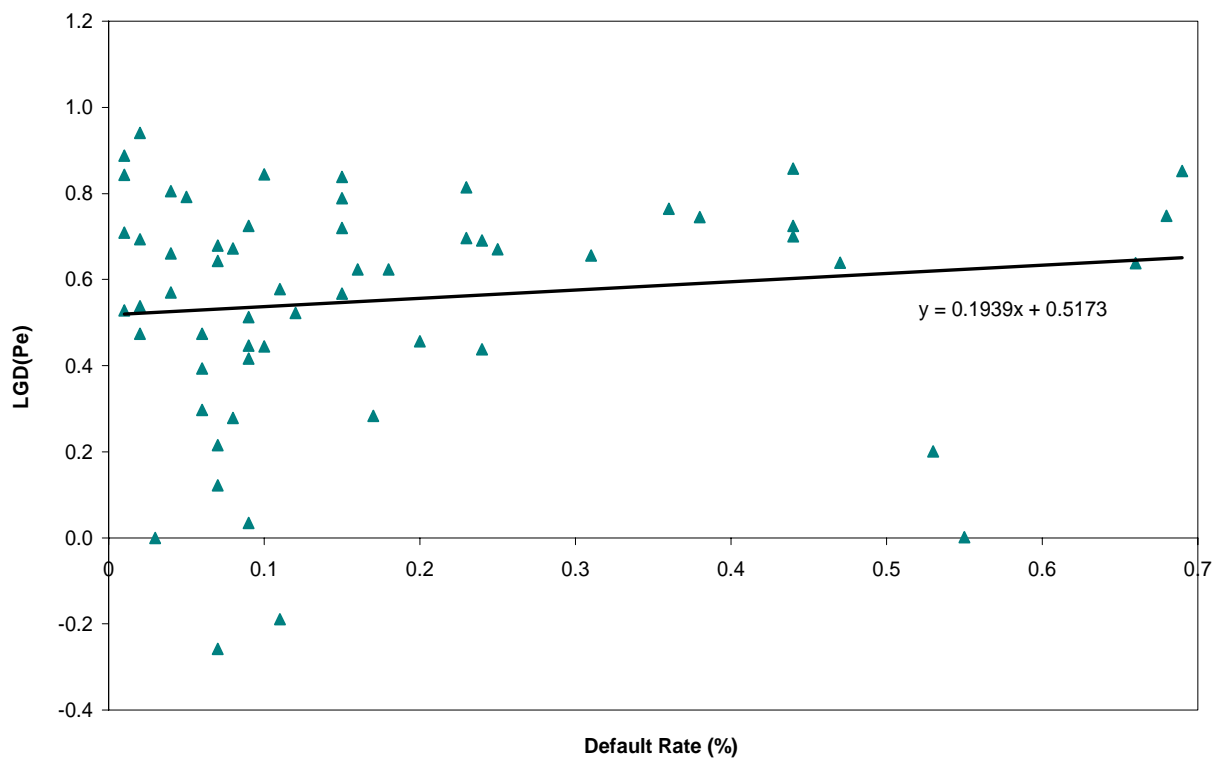


Figure 2 Scatter plot for default rate against LGD( $P_d$ )

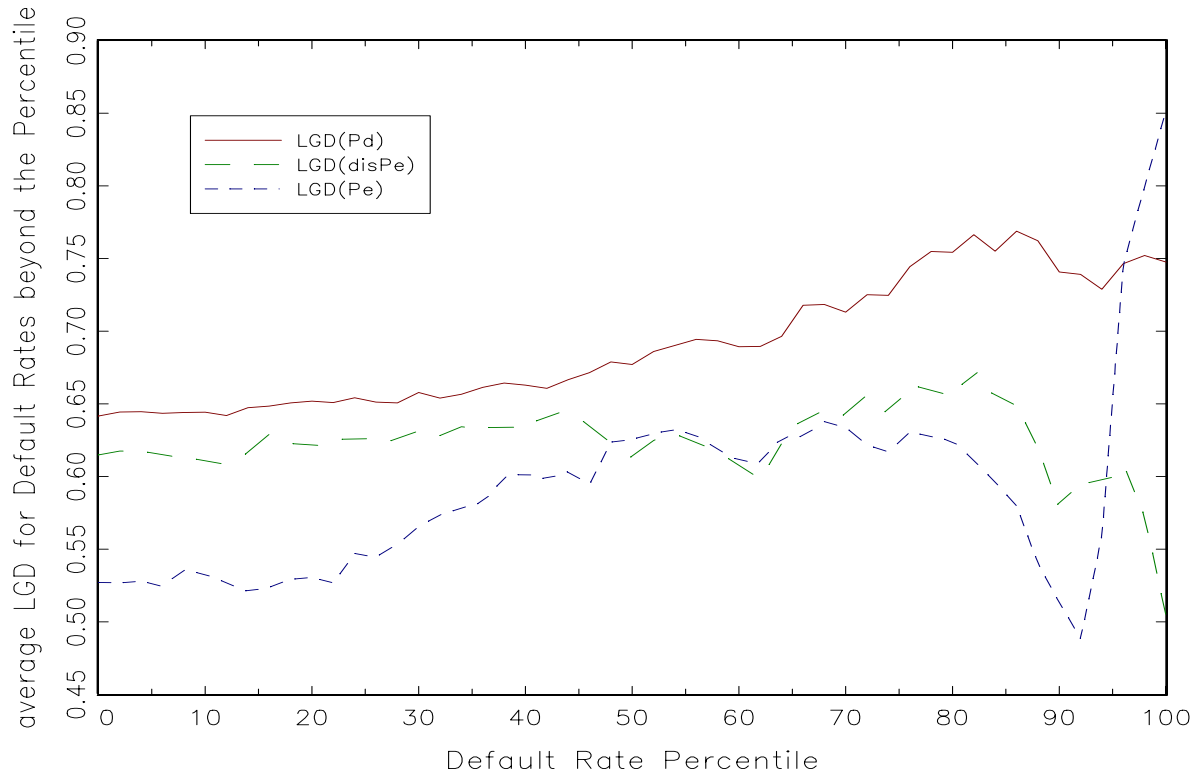


**Figure 3** Scatter plot for default rate against LGD(disP<sub>e</sub>)

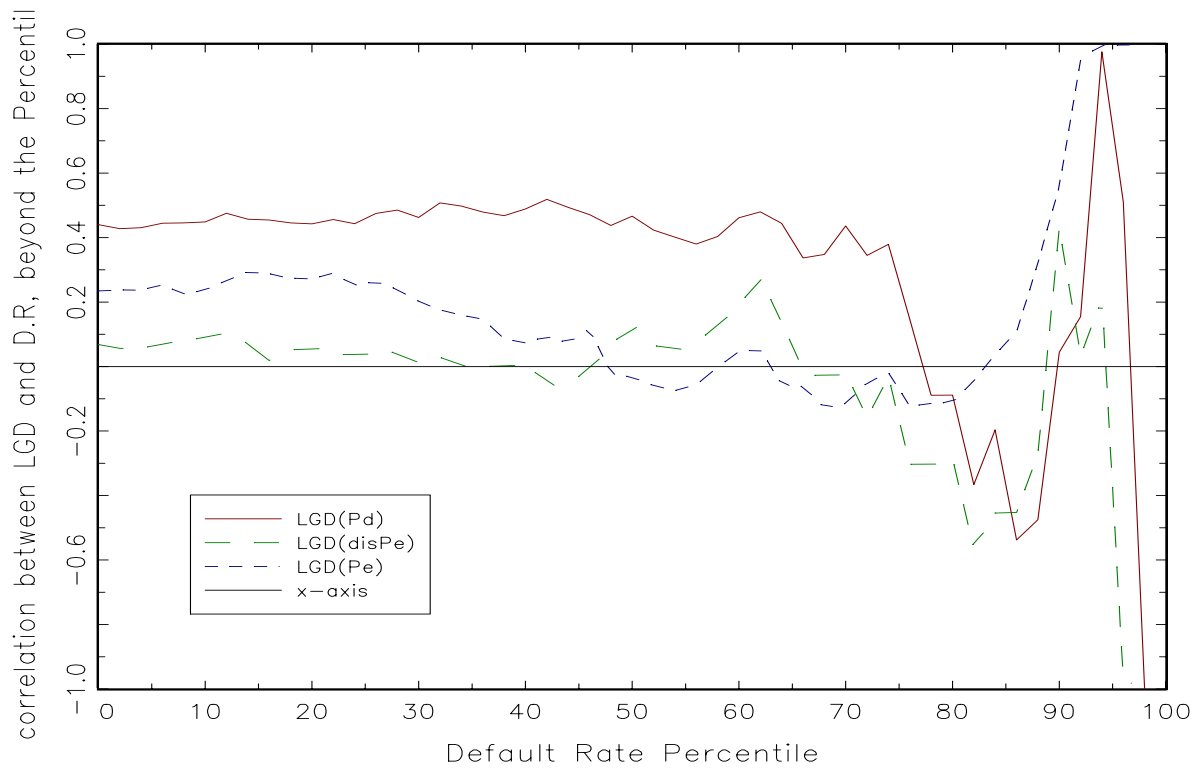


**Figure 4** Scatter plot for default rate against LGD(P<sub>e</sub>)

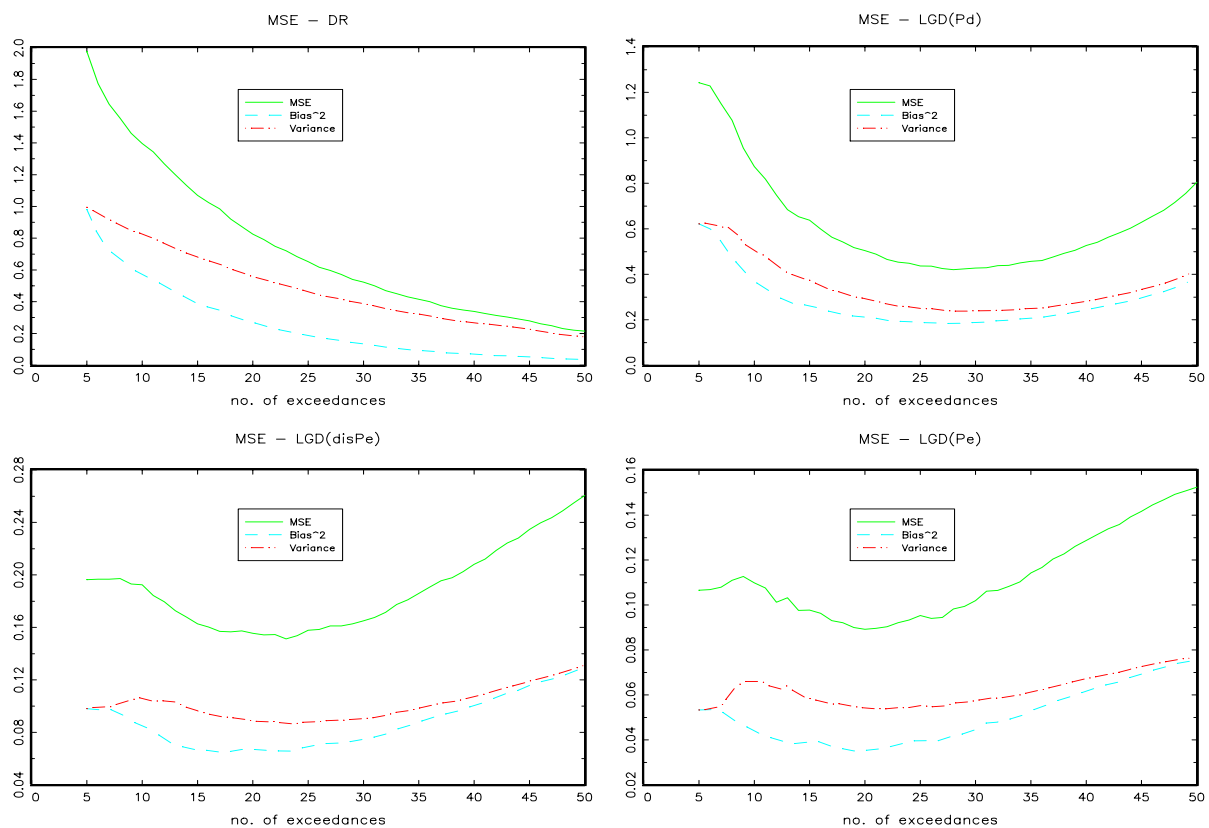




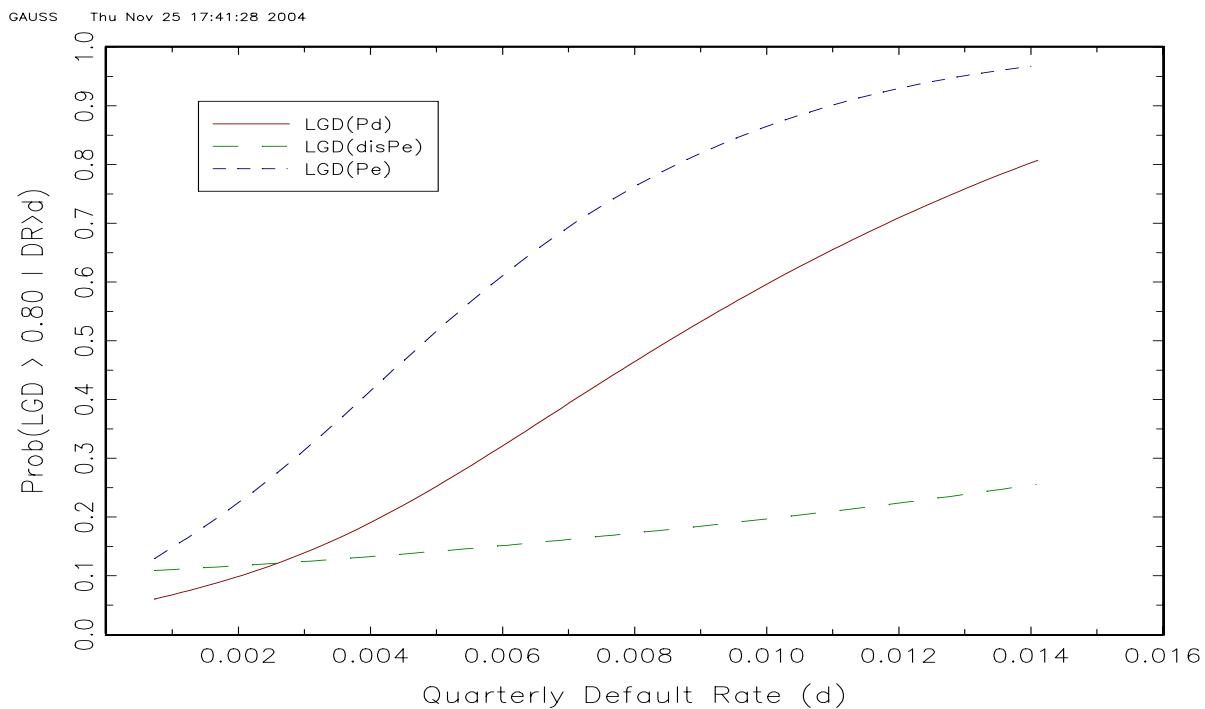
**Figure 5 Average LGD for Default rates beyond the percentile**



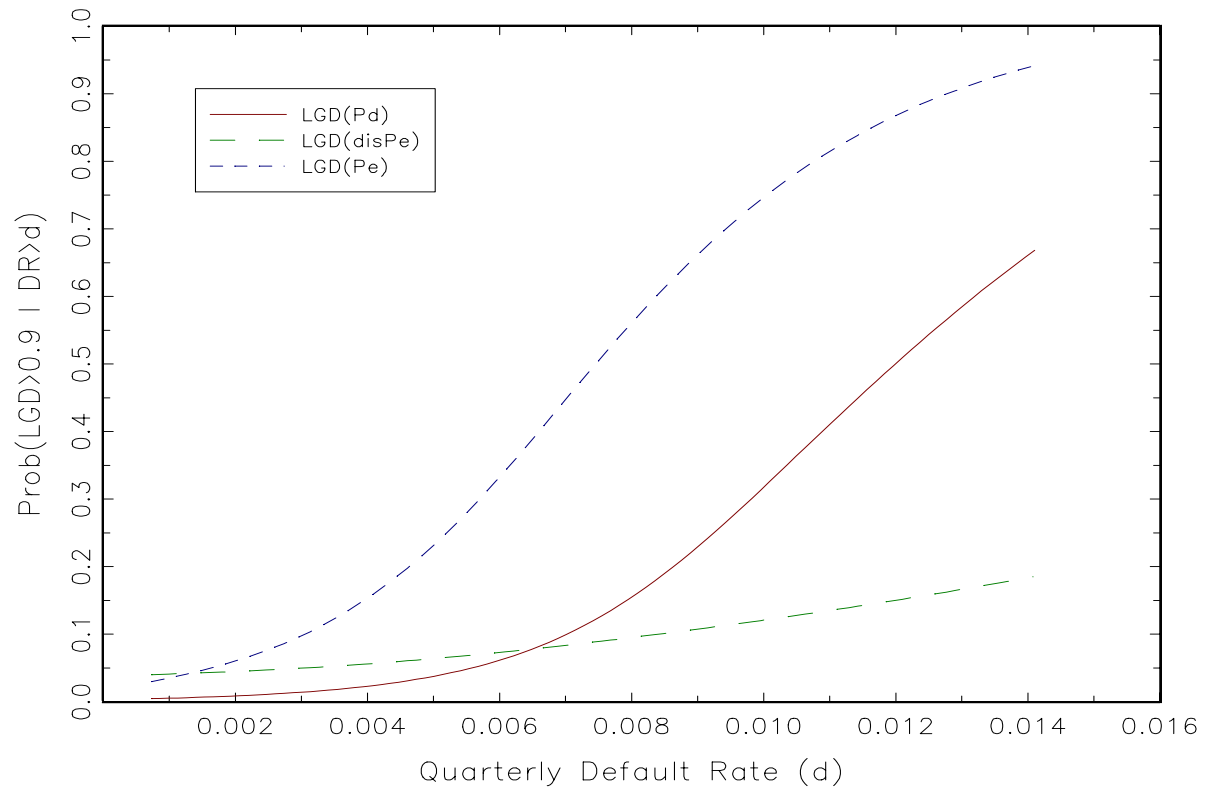
**Figure 6 Conditional Correlation between LGDs and D.R. for D.R. beyond the percentile**



**Figure 7 Mean Squared Error of simulated data with numbers of exceedances from 5 to 50.**



**Figure 8 Conditional Probability of  $\text{LGD} > 0.8$  given a quarterly Default Rate level**



**Figure 9 Conditional Probability of  $\text{LGD} > 0.9$  given a quarterly Default Rate level**